wiiw

Working Paper 248

JUNE 2024

The Granular Trade and Production Activities (GRANTPA) Database

Sebastien Bradley, Javier Flórez Mendoza, Mario Larch and Yoto V. Yotov

The Vienna Institute for International Economic Studies
Wiener Institut für Internationale Wirtschaftsvergleiche

The Granular Trade and Production Activities (GRANTPA) Database

SEBASTIEN BRADLEY
JAVIER FLÓREZ MENDOZA
MARIO LARCH
YOTO V. YOTOV

Sebastien Bradley is Associate Professor at the School of Economics of the LeBow College of Business, Drexel University, Philadelphia. Javier Flórez Mendoza is Economist at The Vienna Institute for International Economic Studies (wiiw). Mario Larch is Professor at the Department of Law, Business & Economics, University of Bayreuth. Yoto V. Yotov is Professor at the School of Economics of the LeBow College of Business, Drexel University, Philadelphia.

We are very grateful to the Eurostat team for research support and clarifications concerning Prodcom codes, combined nomenclature classifications, and complementary concordance files provided by Erlend Firth and Laia Guinovart.

We are grateful to Andrew Bernard, Ilke Van Beveren, and Hylke Vandenbussche for providing an earlier version of this concordance and for sharing the corresponding codes. We also thank David Zenz for his guidance in retrieving the production of manufactured goods database.

This paper expresses the authors' views solely, and Eurostat or any institution is not responsible for any errors.

To access the GRANTPA database, please e-mail us at grantpadatabase@gmail.com.

Abstract

This paper introduces the Granular Trade and Production Activities (GRANTPA) database, which covers international trade flows for 3,124 products and 247 countries over the period 1995-2019 as well as domestic trade flows and production data for the same number of products and years for a subset of 35 European economies. The original data sources that we employ are Eurostat's Comext and Prodcom databases. A gravity application delivers a large set of product-level 'home bias' estimates, which cannot be obtained without domestic trade flows. The average estimates on the standard gravity variables in our model (e.g., distance) are comparable to those from the related literature. However, our disaggregated estimates are very heterogeneous across products, thus highlighting the importance of our new database.

Keywords: Gravity Data, Structural Gravity, Domestic Trade Flows, Disaggregated Gravity Estimates, Home Bias Estimates

JEL classification: C81, F13, F14

CONTENTS

TABLES AND FIGURES

1 Introduction and Motivation

Many important trade policies (e.g., tariffs, export subsidies, etc.), as well as other policies that may impact international trade flows (e.g., technical barriers to trade (TBT), sanitary and phytosanitary standards (SPS), and maximum residue levels (MRL)), are designed and implemented at a very disaggregated level (e.g., 6-digit HS level, which covers more than 5,000 categories as shown in Table 1). Moreover, much of the current policy debate is on the effects of non-discriminatory policies that are country-specific by definition (e.g., TBT, SPS, MRL). However, due to their country-specific nature, the effects of such policies cannot be identified in a properly specified econometric model that only includes international trade flows. A simple theory-consistent solution to identify the effect of country-specific policies on international relative to domestic trade is to rely on domestic (along with international) trade flows (see Heid, Larch and Yotov (2021), Beverelli et al. (2023), and for a survey Yotov (2022)). However, available datasets that include international and domestic trade flows, such as USITC's ITPD-E or CEPII's TradeProd database, are relatively aggregated and do not match the product-level dimension of many trade policies.

To fill this gap, we capitalize on some newly available data from Eurostat and we implement consistent concordance procedures to construct the GRANnular Trade and Production Activities (GRANTPA) database. The sources that we relied on for the construction of the GRANTPA database are Eurostat's *Comext* database, which we used for the bilateral trade data, and Eurostat's *Prodcom* database, which we relied on for the production data. Both of these datasets are collected and maintained by Eurostat. We are indebted to experts at Eurostat, whose guidance was instrumental in using their bulk download facilities to obtain the raw trade and production datasets.

Despite the genuine intent for the European international trade and production classifications (from *Comext* and *Prodcom*, respectively) to be internally consistent over time and also consistent with each other, each of the two databases and corresponding classifications have gone through several changes over time, and many of these changes were specific to each database and independent from each other (both over time and between *Comext* and *Prodcom*). Thus, our first and most demanding task was to construct internally consistent concordances for the international trade data, for the production data, and also between the international trade data and the production data. To this end, we capitalized on and extended previous work by Van Beveren, Bernard and Vandenbussche (2012) and Pierce and Schott (2012*a*,*b*) to construct a new consistent concordance between the raw trade and production data in three steps. First, we ensured the internal consistency of the international trade data (from *Comext*) over time. Then, we made sure that the production data (from *Prodcom*) were also internally consistent over time. Finally, we constructed a concordance between the international trade and production data.

We took four additional steps to construct the GRANTPA database. First, we cleaned and prepared the raw trade and production datasets by eliminating duplicate observations and taking full advantage of the raw data, e.g., by using reported export values to replace corresponding missing import values, etc. Second, we applied our new, consistent concordance to the international trade and production data. Third, we used the bilateral trade data to construct total exports for each product, country, and year in the data, and we combined the total exports with the corresponding production data to construct domestic trade as the difference between the values of production and total exports for each country, product and year in our sample. Finally, we combined the bilateral trade flow data with the domestic trade data to construct the GRANTPA database.

The GRANTPA database covers international trade data for 3,124 products and 247 countries over the period 1995-2019, along with production and domestic trade data for the same number of products and years for 35 European economies, including the 28 EU member states plus potential accessions (Norway, Iceland, Turkey, Montenegro, Bosnia and Herzegovina, Serbia, and Macedonia).[1] To help users who may want to limit their sample and corresponding analysis to only countries for which there is consistent international and domestic trade data, the GRANTPA database includes a 'flag' variable to denote the country-year combinations for which domestic trade data are available.[2]

We demonstrate the usefulness of the GRANTPA database with an application to the workhorse model of trade—the gravity model. Specifically, we obtain estimates of several standard gravity variables including distance, contiguity, common language, and international borders, and we benchmark our results against the large set of existing gravity estimates from the related literature. We draw three main conclusions about the usefulness of the GRANTPA database for gravity analysis. First, the average estimates of the gravity variables that we obtain are comparable to the gravity estimates from the existing literature. Second, while it is possible to obtain gravity estimates of the effects of distance, contiguity, and common language with datasets that only include international trade, our 'home bias' effects can only be obtained with the use of domestic trade flows data, highlighting this important dimension of our data. The home bias estimates that we obtain are large, positive, and statistically significant, which is consistent with the existing literature. However, we are not aware of 'home bias' estimates at such a disaggregated level. Finally, the disaggregated estimates on all gravity variables in our model vary significantly across the products in the GRANTPA database. The implication is that more aggregated gravity analysis may mask significant heterogeneity, which may be important from a policy perspective. Accordingly, we see value in using the GRANTPA database to analyze the effects of various bilateral and country-specific policies.

The rest of the paper is structured as follows. In Section 2, we summarize the procedures that we followed to construct the concordance between the trade and production databases (Subsection 2.1), we describe the raw data and the sources where it was downloaded from (Subsection 2.2), and we highlight the features of the resulting GRANTPA database (Subsection 2.3). Section 3 provides a proof of concept by employing the GRANTPA database to obtain benchmark gravity estimates. Section 4 offers concluding remarks and points to directions for possible uses and improvements of the database. A Supplementary Appendix includes all the detailed steps that we took to construct the GRANTPA database.

[1] Given the period of investigation, the United Kingdom is included as an EU member. Thus, the GRANTPA database features data on international trade flows for a broader set of countries (e.g., the USA and Canada) but does not include data on their domestic trade.

[2] We end up with 3,124 products, i.e., less than the over 5000 6-digit HS categories, because the trade and production data are recorded using different product classification codes that do not fully correspond, i.e., not all product codes in the international trade data have a correspondence to product codes in the production data and vice versa.

2 Sources and Methods

This section describes the methods that we developed to construct the GRANTPA database, the sources for the raw data that we used, and the main features and possible limitations of the GRANTPA database. The section consists of 3 subsections. In Subsection 2.1, we describe the procedures that we followed to construct internally consistent concordances for the international trade data, for the production data, and also between the international trade data and the production data. In Subsection 2.2, we summarize the sources where the raw trade and production data were accessed and downloaded from. Finally, in Subsection 2.3, we deploy the raw trade and production data that we describe in Subsection 2.2 and we use the concordances that we created in Subsection 2.1 to construct the GRANTPA database. Specifically, in this subsection we describe the additional steps that we took to prepare each of the raw databases for merging and we combine the trade and production data to construct domestic trade. Then, we combine the international and domestic trade data to deliver the final version of the GRANTPA database. Subsection 2.3 concludes with a description of the main features, dimensions, and limitations of the GRANTPA database. For a very detailed description of the data sources and the steps that we took to construct the GRANTPA database, we refer the reader to the Supplementary Appendix.

2.1 Concordances for the Trade and Production Data

Despite the genuine intent for the European international trade and production classifications (from *Comext* and *Prodcom*, respectively) to be internally consistent over time and also consistent with each other, each of the two databases and corresponding classifications have gone through several changes over time, and many of these changes were specific to each database and independent from each other (both over time and between *Comext* and *Prodcom*). As a result, the most demanding task in the creation of the GRANTPA database was to construct a consistent concordance between the international trade and production datasets.

To this end, and following previous studies, we constructed the needed concordance in three broad steps, which correspond to the three subsections of this section. Specifically, in Subsection 2.1.1, we describe the procedures that we followed to ensure internal consistency of the international trade data (from *Comext*) over time. Then, in Subsection 2.1.2, we summarize the steps that we took to make sure that the production data (from *Prodcom*) is internally consistent over time. Finally, in Subsection 2.1.3, we describe the procedures to construct a concordance between the international trade and production data. Each subsection starts with a summary of the challenges that we faced, followed by a summary description of the methods that we applied to address them. The Supplementary Appendix includes a detailed description of the procedures that we describe here.

We benefited from and expanded upon the methods from several related efforts at each of the three steps to create the data concordances. Specifically, Van Beveren, Bernard and Vandenbussche (2012) (henceforth, VBBV) focus on the implications of changing product classifications using Belgian firms. Even though our focus is broader (i.e., we aim to construct a consistent trade and production database for Europe), we benefited tremendously from the guidance and concordances that were created by

VBBV.[3] In addition, we also implemented some of the algorithms provided by Pierce and Schott (2012*a*,*b*) for concording the US Harmonized System codes over time and with the SIC/NAICS product classes and industries. Finally, since the GRANTPA database includes many new/additional years (e.g., the concordances from VBBV end in 2010, whereas ours run through 2022) during which the underlying international and production databases and classifications have changed significantly, we had to address some new challenges, and we utilized new concordance files extracted from Eurostat.

2.1.1 Concording the International Trade Data Over Time

The European international trade data at the product level from *Comext* is recorded according to the *8-digit Combined Nomenclature* (CN8) classification. Due to various changes in the product classification and the composition of products in the trade data (e.g., some products disappear while new products emerge),[4] to have a consistent international trade data, we need to construct a common and consistent concordance for the products over time, which is labeled CN8+. To this end, and following VBBV,[5] We proceed in three steps.

- ***Step 1: Concording over consecutive years***: Most CN8 products remain unchanged between two years, i.e., they are mapped one-to-one. However, there is also a need to concord the CN8 classification over consecutive years for one of several reasons, including disappearing products, new products, split of one product category in one year into several product categories in a consecutive year (one-to-many), combining several product categories in one year into a single category in a consecutive year (many-to-one), and mapping many-to-many products. Therefore, the first step to ensure internal consistency over time is to create a variable that identifies each 'family' of codes, i.e., codes that are connected over consecutive years. For the many-to-many and one-to-many mappings between two years, we rely on a 'feedback' loop from Pierce and Schott (2012*b*), which creates a synthetic code that keeps track of codes that need to be grouped in "*family trees*."[6] Table 2 reports the number of obsolete and new codes in each year, the number of families, and the number of simple changes for the years in the trade data.

- ***Step 2: Concording over the whole period***: An additional challenge arises because some product codes change in more than one year. This requires a procedure that ensures consistency over the whole coverage period of the database. To this end, we take three steps. First, we again rely on the so-called "news loop" developed by Pierce and Schott (2012*b*) to link the codes that underwent multiple

[3]The concordances from VBBV are available at: https://sites.google.com/site/ilkevanbeveren/concordances.

[4]Table 1 illustrates the evolution of the 8-digit combined nomenclature (CN8) classifications until 2022.

[5]For consistency in the terminology, we use various labels and variable names from VBBV who, in turn, follow Pierce and Schott (2012*a*,*b*).

[6]Since synthetic codes group original products that were recorded either in a more detailed (shrinking family) or less detailed (growing family) manner in previous years, the number of synthetic codes and hence the level of detail of the final (synthetic) product classification will be reduced as the time period under consideration grows longer. This also implies that, depending on the beginning and end year of the concordance, the family trees can be different.

changes over time into "families/chains."[7] For each new code in a particular year, the algorithm searches for matching (identical) obsolete codes in later years. If a new code has become obsolete in later years, the two families of which the code is a part are chained together. These family trees can then be merged back into the file with all obsolete and new mappings. Second, these chains/families are superimposed on the year-by-year concordance from the previous step. Finally, these families are merged into the full list of existing CN8 codes in each year to translate the CN8 codes into the CN8+ classification.

Moreover, our approach entails aligning identified "chains" onto the year-by-year concordance. The foundation of this process lies in "loop 1" in VBBV (*generating set years for changes between t and t-1*), where we assign unique identifiers to each code change, even in complex mappings (one-to-many and many-to-one). The essence of this loop ensures that every code change, no matter how complex the mapping might be, receives a unique identifier. This alignment allows us to trace code changes accurately and consistently within each specific year of the dataset. In addition, we implement the "news loop", which extends to a broader scale. Here, we integrate these detailed code families into the entire list of CN8 codes. This integration forms the enriched CN8+ classification, which includes all codes, encompassing both those that have remained constant and those that have changed over time. The "news loop" finds "news" or updates from subsequent years that inform or modify "news" from earlier years. The "news loop" essentially weaves together a comprehensive process detailing how a code evolves over multiple years. By combining "loop 1" and the "news loop", we ensure a robust and meticulous merge, balancing the yearly changes with the completeness of the CN8+ classification.

- ***Step 3: Constructing consistent international trade data***: The last step in the construction of the international trade data is to merge the concordance between CN8 and CN8+ with the trade data. In doing so, two considerations were taken into account. First, since, by construction, the concordance between CN8 and CN8+ is time-varying, two identifier variables are required to merge the concordance with the trade data: namely, 'year' and 'CN8' identifiers. Second, due to the existence of 'many-to-one', 'one-to-many', and 'many-to-many' cases, an aggregation/collapsing of the data is necessary to end up with a unique code-year observation for each country pair.

The resulting international trade database is used for the construction of domestic trade values (in combination with the production data, whose construction we describe next) and enters directly into the final GRANTPA database. The international trade database includes the following four variables for this purpose: (i) "*year*", which is a numeric variable denoting the year of trade; (ii) "*cn8*", which is a unique, year-specific, numeric CN8 code; (iii) "*synthetic*", which is a dummy variable equal to one if the CN8+ classification groups more than one CN8 code; and (iv) "*cn8plus*", which is a unique, year-specific, numeric CN8+ code. Finally, we note that the original trade data, as well as the consistently concorded version of the international trade data that will

[7]Specifically, this loop identifies families of codes by searching for updates of new codes in later years, and it only retains codes that have undergone multiple changes over the time period considered; hence the family trees have to be combined with the original mappings to obtain a final unique identifier that keeps track of changes between two years and family trees over time.

be included in the GRANTPA database cover trade between all European countries as well as trade between all European countries and non-European partners. Production data are not available for non-European countries; nevertheless, we retain trade values between European and non-European economies in the GRANTPA database to render the database as useable as possible for different applications.

2.1.2 Concording the Production Data Over Time

Prodcom ("*Production Communautaire*") is a system used in the EU to compile statistics on the production of manufactured goods in member states. It includes both a database and a product classification system. Companies are required to report their industrial production and services using the *Prodcom* list. They must record their production activities at the 8-digit *Prodcom* level (PC8) on a monthly basis. The *Prodcom* declaration includes data on the physical volume and value of production sold for each product during the survey period.

Following the insights from VBBV, we note several limitations and challenges when using the *Prodcom* data:

- ***Inconsistency in Coverage***: The coverage of the CN8 classification remains constant across years, whereas the *Prodcom* list changes over time. This inconsistency means that a product may be covered by a *Prodcom* code in one year but not in another. As a result, production in these codes cannot be tracked consistently over time, leading to the exclusion of such data in time-based concordances. To address this challenge, production data that cannot be consistently tracked over time due to changes in *Prodcom* codes are excluded from time-based concordances.

- ***Classification Variability***: The *Prodcom* classification system includes optional B-list and N-list codes, which are not uniformly used across countries. This variability complicates the calculation of EU totals for these products. Thus, we focused solely on mandatory 8-digit *Prodcom* codes, excluding optional and aggregated codes from our analysis, except where these optional codes affect mandatory codes.

- ***Existence of Aggregated Prodcom Codes***: *Prodcom* tracks different types of aggregated codes (such as Q-, V-, Z-, T-, and E-list), which makes it challenging to analyze data consistently and in detail. To avoid the issue of double counting, we implement a concordance procedure that flags aggregate codes and drops the corresponding disaggregated codes to achieve consistency. This procedure considers changes in optional and aggregated codes that only affect mandatory codes and ensures consistent treatment and detailed analysis of aggregated codes.

- ***Alignment between CPA6 and NACE 4 Classifications***: The first six digits of *Prodcom* codes correspond to CPA6 products, which are classified according to the Classification of Products by Activity (CPA6). Nevertheless, not all CPA6 codes are covered by the *Prodcom* list. To ensure alignment and consistency, it is necessary to align with the NACE 4 classification for economic activity categorization. The concordance procedure fully controls and treats PC8 codes in relation to CPA6 classifications by focusing on mandatory PC8 codes and their correspondence.

We develop a concordance procedure to thoroughly address each of the foregoing *Prodcom* data challenges. This enables us to account for inconsistencies, variability, and alignment with CPA6 and NACE 4 classifications. Our approach focuses exclusively on mandatory 8-digit codes and covers the period 1995 to 2019. Recoding of obsolete and new mandatory codes allows us to track changes over time. As part of this procedure, we also recode optional codes into mandatory codes per guidelines in VBBV. This systematic approach ensures the reliability and validity of our analysis.

- ***Step 1: Concording PC8 Codes Between t and t-1***: The first step involves adapting to changes in PC8 codes year-on-year. We categorize these changes as simple mappings: one-to-one, many-to-one, one-to-many, and many-to-many. A unique identifier 'setyr' is used to track these mappings. For complex mappings like many-to-many and one-to-many, we apply a feedback loop based on Pierce and Schott (2012*b*) to create synthetic codes and form 'family trees' of connected codes.
- ***Step 2: Developing a Consistent Classification Over Time***: In the second step, we address the challenge of codes changing in more than one year. Using the "news loop" technique from Pierce and Schott (2012*a*,*b*), we link these changes to form chains or 'families' of codes, ensuring consistency over the database's entire time period. This results in a unique identifier that captures both the year-to-year changes and the overarching family trees of codes.
- ***Step 3: Concording Production Data***: The final step involves aligning the European domestic production data with the refined PC8+ classification. Here, optional codes are recoded into their mandatory counterparts for consistency. The main focus is on the *year* and *pc8* variables, ensuring that each product code and year combination yields a unique observation. The result is a dataset with key variables including *year*, *pc8*, *synthetic* (indicating if PC8+ groups more than one PC8 code), and *pc8plus* (the year-specific PC8+ code).

2.1.3 Concordance Between the Trade and Production Data

To concord international trade and domestic production data at the 8-digit product level, a common classification, PC8+, is used to bridge CN8 product codes (for international trade) and PC8 codes (for domestic production). Nonetheless, concording trade and production data within a single year presents several challenges due to differences in coverage between the PC8 and CN8 classifications.

- ***Coverage Discrepancies Between CN8 and PC8***: Not all CN8 products are covered by the *Prodcom* list. This discrepancy requires the exclusion of specific CN8 codes from the international trade data, as these codes do not appear in the PC8-CN8 concordance. To address this challenge we used year-specific lists to identify CN8 codes not covered by *Prodcom* and exclude them from the international trade data. The remaining CN8 codes can then be translated into the PC8+ classification.
- ***Incomplete Coverage of PC8 by CN8***: Certain PC8 products, for instance, industrial services and waste products, are not covered by the CN8 classification. This includes disaggregated codes of aggregated PC8 products. The exclusion

of industrial services and recoding of optional and disaggregated PC8 products into their mandatory and aggregate counterparts enable us to concord the PC8 classification into the PC8+ classification.

- ***Combining Data Over Time***: When combining data on international trade and domestic production over time, the changing coverage of the *Prodcom* list and the difference in coverage between PC8 and CN8 classifications must be considered. The methodology considers changes in the *Prodcom* list coverage and the differences and updates in CN8 and PC8 classification systems over time.

Following the methodology outlined by VBBV, the concordance procedure is implemented in four steps. The first step deals with concording product classifications within a single year. The remaining steps focus on the actual implementation of these concordances in the international trade and production data, ensuring consistency and accuracy in the alignment between *Comext* and *Prodcom*.

- ***Step 1: Concordance from CN8 and PC8 to PC8+***: This step involves aligning all PC8 codes covered by CN8 classification into PC8+ products. Mappings between CN8 and PC8 include various types, and a unique identifier 'setyr' is used for each mapping. A feedback loop, based on Pierce and Schott (2012*b*), ensures correct groupings for complex mappings.[8] The outcome is a concordance file that lists CN8 codes, their corresponding PC8 codes, and the assigned PC8+ product code.

- ***Step 2: Concording the Production Data***: For concording European 8-digit production data to PC8+ products, steps include recoding optional codes into mandatory counterparts and dropping PC8 products not covered by CN8. Disaggregated PC8 codes are recoded into aggregated codes where necessary. The process ensures that all relevant PC8 codes are included in the concordance, with a focus on achieving a unique product code.

- ***Step 3: Concording International Trade Data***: The international trade data is concorded by merging it with the concordance file, ensuring each CN8 code is represented once, with its corresponding PC8+ product code. CN8 products not covered by PC8 are excluded. The procedure ensures uniqueness in the trade data, with an emphasis on avoiding double-counting.

- ***Step 4: Merging Domestic Production and Trade Data***: The final step involves sorting and merging the domestic production and international trade data based on product identifiers. The merged data set uses the PC8+ classification, allowing for product-level comparisons between trade and production data as shown in Figure 5.

In addition to these steps, specific do-files are created for each year, ranging from 1995 to 2019, to facilitate the concordance between classifications.[9] These do-files generate essential input files required for the concordance process, ensuring that all necessary steps are taken to align domestic production and trade data in a common classification.

[8]Figure 2 shows an example of the one-to-many codes in 2003. Similarly, Figure 3 displays the many-to-one codes in the same year. Lastly, Figure 4 represents the many-to-many codes.

[9]Year-specific do-files are crucial as each year in the raw files varies in format and coding systems. These files adapt to classification descriptions, optional, aggregate, and special PC8 codes, ensuring consistent data cleaning and alignment across the period of analysis.

2.2 Raw Data and Sources

In this section, we briefly describe the raw trade and production data and the sources that we used to retrieve it. We provide further details in the Supplementary Appendix. The original database for the international trade data we used to construct the GRANTPA database is *Comext*, which we extracted using Eurostat's bulk download facility. The international trade data are split into two periods. Data for the period 1988-2001 are located at *Comext* Historical Data and can be downloaded using the following path: *Comext files* > *COMEXT_HISTORICAL_DATA* > PRODUCTS_1988_2001.

Data for the period 2002-2022 are accessible via *Comext files* > *COMEXT_DATA* > PRODUCTS.[10] *Comext* has several dimensions, which we capitalize on in the construction of the GRANTPA database. First, it records international trade values for countries within the EU as well as between EU and non-EU member countries. The group of destination and origin countries in the intra-EU and extra-EU declaration has changed over time due to changes in EU membership.[11] In addition, *Comext* includes two flows for each pair of countries, e.g., exports from Austria to Bulgaria and imports in Bulgaria from Austria. We utilize this feature by replacing some missing values in one direction of the trade flows with the corresponding values in the other direction, which are not missing. Finally, *Comext* reports data on trade values and quantities. For the GRANTPA database, we use the values of trade. However, as described in the next section, we also use trade quantities to distinguish between missing values versus true zero trade values in the data.

Production data are only available for European countries. These data are downloadable from Eurostat's bulk download facility as '*Europroms*'.

Due to changes in classifications over time, the production data must be downloaded separately for different periods. We describe the specific download sources and the steps undertaken to obtain the raw production data in more detail in the Supplementary Appendix.

Similar to *Comext*, we capitalize on several dimensions of the *Prodcom* database. Specifically, in addition to the values of production at the product level, *Prodcom* reports quantities, which we use to distinguish between true zeroes versus missing production values when the latter values are not reported. In addition, *Prodcom* classifies some values as confidential (C:). These must necessarily be treated as missing values. Finally, *Prodcom* includes data on the value of total exports, which we use to construct domestic trade flows—i.e., the difference between total production and exports—for each country, year, and product in our sample.

2.3 The GRANTPA Database: Construction and Coverage

In this subsection, we deploy the raw trade and production data that we described in Subsection 2.2, and we use the concordances that we created in Subsection 2.1 to construct the GRANTPA database. We proceed in 5 steps. We start with a description

[10] The original data is retrieved on the date 26 December 2022.

[11] In 1995, Austria, Finland, and Sweden joined the EU. Then, in May 2004, ten new countries joined: Cyprus, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Malta, Poland, Slovakia, and Slovenia. In 2007, Bulgaria and Romania joined. Lastly, Croatia joined the EU in 2013 and more recently the United Kingdom exited the EU in 2019. When a country joins the EU, its exports and imports vis-à-vis EU countries are no longer reported in the extra-EU trade statistics. Instead, they have to be declared in the intra-EU trade statistics.

of the additional steps that we took to prepare the raw trade database (Step 1) and raw production database (Step 2) for merging them with each other. Then, in Step 3, we apply the new concordances to the trade and production databases. In Step 4, we combine total product-level exports and production values for each country and year in our sample to construct domestic trade. Finally, in Step 5, we combine the international and domestic trade data to construct the GRANTPA database. We conclude with a description of the main features, dimensions, and limitations of the GRANTPA database, and we flag three alternative subsets of the data based on country coverage. The Supplementary Appendix provides a more detailed description of these steps.

- ***Step 1: Prepare the international trade data.*** As discussed in the previous section, the international trade data from *Comext* has several dimensions (e.g., intra- versus extra-EU trade and imports versus exports). We try to take full advantage of the information that is contained in the raw data.

  First, we eliminate any duplicate observations, which could appear due to double reporting or because of the letter codes (T-,Q-,V-,E-) that are included in the raw trade data. We focus on import values from *Comext* as our main trade variable. Thus, if *Comext* provides data on both exports from Germany to France and imports in France from Germany for a particular product, we use the reported imports (in France from Germany). The reason for selecting imports is that, by definition, the reported import values include trade costs (e.g., cost, insurance, and freight). This is consistent with trade/gravity theory, which is derived at delivered prices.

  Even though we rely on imports as our main variable, we also take advantage of some of the information contained in the reported export values. For example, if data on product-level imports for a given pair are not reported or zero, we replace these import values with their trading partner's corresponding (non-zero or non-missing) exports. Finally, we used the *Comext* trade data to create total exports, which, as described below, are needed to construct domestic trade. To construct total exports, we use positive bilateral import data to replace corresponding missing or zero bilateral export values. Next, we sum all bilateral exports for each year and country. Importantly, we utilize the fact that *Comext* includes trade between each of the European economies and all other countries in the world.

  Processing the *Comext* data in this manner yields two key variables: (i) an unbalanced product-level bilateral trade variable for trade between each of the European countries and all other countries in the world, including the European countries themselves, and (ii) an unbalanced total exports variable for each European country (by product-year).

- ***Step 2: Prepare the production data.*** We need the production data (in combination with total exports) to calculate domestic trade flows. To construct product-level production for the years and countries in our sample, we take advantage of several dimensions of the *Prodcom* database. Similar to the trade data, we first eliminate duplicate production-value observations, which could appear due to double reporting or because of the letter codes (T-,Q-,V-,E-) that are included in the raw production data. Then, we select the reported product-level values of production for each country and year as our main production value variable. We take advantage of the fact that *Prodcom* reports quantities as well as values by making sure that we treat missing production values as true missing values (as opposed to zeros) when the corresponding reported quantities are not missing. Finally, we also make sure that we treat missing production values as true missing values when *Prodcom* includes a corresponding code for confidential data. The outcome of this step is an unbalanced production value variable for each product, country, and year in our sample.

- ***Step 3: Add concordances to the trade and production data.*** We use the concordances created in the previous step to make sure that the trade and production data are consistently classified. We apply the new concordances to the bilateral trade data, the total exports data, and the production values data. As expected, due to the presence of one-to-many, many-to-one, and many-to-many combinations, we have to aggregate some of the product categories in each of the datasets, so that the product classification is unique and common across the trade and production datasets.

- ***Step 4: Construct domestic trade.*** In this step we combine the data on total product-level exports and product-level production values for each country and year in our sample, and we construct domestic sales as the difference between total production and total exports. Then, we replace any resulting negative values with missings. Similar to Bernard et al. (2018), we observed cases where exports exceed production. While the exact values are year-specific, the percentage of negative values ranges between 9% in 1995 to 22% in 2019. As proposed by Bernard et al. (2018), 'Carry Along Trade' is a possible explanation. Another reason is changes in inventories.

- ***Step 5: Construct the GRANTPA database.*** In this step, we construct the full and final version of the GRANTPA database. To this end, we combine/append the bilateral trade data that we constructed in Step 3 with the domestic trade data that we constructed in Step 4. The result is the GRANTPA database, which covers 247 countries and 3,124 products over the period 1995-2019. Table 4 of the Supplementary Appendix includes the list of all countries in the GRANTPA database, while a list of the products that are covered in the GRANTPA database, along with the corresponding codes from *Comext* and *Prodcom* can be found here.

  The GRANTPA database covers bilateral trade between 35 European economies and all other countries in the world. However, it should be noted that the GRANTPA database (i) does not include trade between non-European countries, and (ii) the domestic trade data are available exclusively for European countries for which we have total trade and production data, i.e., no domestic trade data are available for non-European economies, and domestic trade data may even be missing for some European countries (e.g., late EU joiners). To help users who may want to limit their sample and corresponding analysis to only countries for which there is consistent international and domestic trade data, the GRANTPA database includes a 'flag' variable to denote the country-year combinations for which there is domestic trade data.

  The GRANTPA database includes the following variables: (i) *'exporter'* is a 3-letter ISO code for each exporter;[12] (ii) *'importer'* is a 3-letter ISO code for each importer;

[12] See Table 4 of the Supplementary Appendix for a list of all countries in the GRANTPA database and their full names.

(iii) *'year'* captures the year when trade took place; and (iv) *'product'* is a numeric product code which is unique to the GRANTPA database. A complete concordance file, which includes product IDs, product names, and the corresponding codes from *Comext* and *Prodcom* can be found here. (v) *'trade'* denotes nominal (domestic and international) trade values in thousands of Euros.[13] Finally, (vi) *'flag'* is an indicator variable that takes a value of one for the countries (and years) for which the database includes domestic trade data and zero otherwise.

3 Gravity with the GRANTPA Database: A Proof of Concept

The objective of this section is to deploy our new GRANTPA database in an application as a proof of concept. To this end, we selected a 'gravity' application for two main reasons. First, the main motivation for constructing the GRANTPA database was that it could be employed for disaggregated gravity analysis at the product level. Second, the gravity model is the workhorse model of trade and, as such, it has been employed in thousands of papers that study various determinants of trade flows. Thus, we can rely on a large set of existing gravity estimates against which we can benchmark our new results to establish the representativeness and credibility of the GRANTPA database. We proceed in three steps. First, we combine the GRANTPA database with some existing gravity datasets. Then, we specify our estimating gravity model. Finally, we obtain and interpret our results.

The first gravity database that we combine with our GRANTPA database is the US International Trade Commission's *Dynamic Gravity Database* (DGD), which is created and maintained by Gurevich and Herman (2018). We use the DGD to obtain the covariates for bilateral distance and contiguity. In addition, we rely on the *The Domestic and International Common Language* (DICL) database of Gurevich et al. (2023) to obtain a variable for common language. We rely on the DICL dataset because it includes a continuous variable for common international language, which, as demonstrated by Gurevich et al. (2021), dominates the use of a dummy variable for common language, which is the standard approach in the literature. Finally, we construct a dummy variable that takes a value of one for domestic trade and is equal to zero otherwise. The estimates of this variable will reflect the effects of forces that drive a wedge between domestic and international trade, which are referred to in the literature as 'home bias' effects. Identifying such effects is not possible without the availability of domestic trade data—one of the core attributes of the GRANTPA database and one of the main motivations for its construction.

Capitalizing on some of the current gravity estimation techniques, as summarized by Yotov et al. (2016), we specify the following simple gravity model:

$$\begin{aligned} X_{ij,t}^{k} \;=\; & \exp[\gamma_1^k DIST_{ij} + \gamma_2^k CNTG_{ij} + \gamma_3^k LANG_{ij}] \times \\ & \exp[\gamma_4^k SMCTRY_{ij} + \psi_{i,t}^k + \phi_{j,t}^k] \times \varepsilon_{ij,t}^k, \qquad \forall i, j. \qquad (1) \end{aligned}$$

Here, $X_{ij,t}^{k}$ denotes the nominal exports (at delivered prices) of product k from exporter

[13] We note that, by construction, if we sum the values of trade (including domestic trade) for each exporter and year for which the GRANTPA database includes domestic trade, we will obtain the corresponding production values for each product, country, and year.

i to destination j at time t.[14] Following Santos Silva and Tenreyro (2006), we estimate equation (1) using the PPML estimator,[15] which accounts for potential heteroskedasticity issues inherent to trade data and enables us to take advantage of the information that is contained in the zero trade flows in the GRANTPA database. The gravity covariates in equation (1) include the logarithm of the bilateral distance between the trading partners ($DIST_{ij}$) and indicator variables for the presence of contiguous borders ($CNTG_{ij}$), common official language ($LANG_{ij}$), and domestic vs. international trade ($SMCTRY_{ij}$). Finally, following the literature, we cluster the standard errors by country-pair.

We rely on specification (1) to obtain a set of gravity estimates for each of the 3,124 products in the GRANTPA database. Due to the large number of estimates, we report them, along with their corresponding confidence intervals, in Figure 1. For clarity of exposition (due to the presence of outliers), we do not include the largest and smallest five percent of point estimates for each of the gravity variables in our model. In addition, we drop the top and bottom five product-level estimates with the widest confidence intervals. The four panels of Figure 1 report the estimates for each of the four gravity variables in our model, and in each case, we have ordered them from smallest to largest.

Panel A of Figure 1 reports the results for distance—the most widely used and robust gravity covariate. The main conclusions that we draw from this figure are threefold. First, most of the estimates (about 94%) of the effects of distance on product-level trade are negative and statistically significant, which is consistent with the voluminous gravity literature. Second, in terms of magnitude, the average of the distance estimates is -0.769 (std.dev. 0.618), which is also readily comparable with the vast majority of the distance estimates from the existing literature. Third, the estimates of the effects of distance are quite heterogeneous across the products covered by the GRANTPA database. This is important for the current purposes because the wide variation in the estimates of the distance effects that we obtain suggests that more aggregate gravity estimates mask significant heterogeneity, which may be very important from a policy perspective.

Without going into too much detail, we note that the estimates on contiguity and common international language are both mostly positive and statistically significant. Specifically, 78% of the estimates of the effects of contiguity that we obtain are positive and most of them are statistically significant. Similarly, 80% of the estimates of the effects of common language are positive and, once again, most of them are statistically significant. These results are also consistent with findings from the existing literature and imply that sharing a common border and speaking the same language promote international trade. In terms of magnitude, the average estimates on common borders (0.316, std.dev. 1.174) and common language (0.876, std.dev. 2.572) are also very similar to corresponding estimates from the existing literature. In addition, we also observe very heterogeneous estimates for these two variables, thus reinforcing the argument for using disaggregated data for gravity estimations.

Finally, we turn to the estimates on the $SMCTRY$ variable, which are reported in Panel D of Figure 1. Importantly, these estimates can only be identified due to

[14]The disaggregated estimating gravity equation (1) has solid theoretical foundations on the demand side, e.g., Anderson and van Wincoop (2004), and on the supply side, e.g., Shikher (2011) and Costinot, Donaldson and Komunjer (2012). Following Arkolakis, Costinot and Rodríguez-Clare (2012), Yotov et al. (2016) employ the notation of Anderson and van Wincoop (2003) to demonstrate the equivalence between the industry-level gravity equations on the demand side and the supply side and discuss the implications for gravity estimations.

[15]In practice, we use the fast and robust estimation command 'ppmlhdfe' of Correia, Guimarães and Zylkin (2020).

the domestic trade dimension of the GRANTPA database. As expected, most of the $SMCTRY$ estimates (more than 90%) that we obtained are positive, and most of them are statistically significant. This result, sometimes dubbed as the 'home bias' effect, is well-established in the gravity literature and reflects the fact that *ceteris paribus*, most sales are domestic. What is novel, however, is that for the first time in the literature, we confirm this result with very disaggregated data. In terms of magnitude, the average estimate on $SMCTRY$ that we obtain is 1.741 (std.dev. 1.513), and it implies that *ceteris paribus* domestic trade is about 4-5 times larger than international trade. We find this implication plausible, and it is comparable to recent estimates from the gravity literature.

Finally, and similar to the estimates on the other gravity variables, we observe very wide heterogeneity in the 'home bias' effects at the product level. We believe that exploring this heterogeneity further, e.g., investigating its drivers or variation across countries, etc., could be very interesting and important from a policy perspective. Similarly, we know that our gravity specification can be improved and extended to include several other important determinants of trade flows, e.g., various bilateral as well as country-specific trade policies. However, since our current purposes are simply to demonstrate the usefulness and applicability of the GRANTPA database for gravity estimations, we leave this type of more detailed analysis for future work.

4 Conclusion

This paper introduced *The Granular Trade and Production Activities* (GRANTPA) database, which covers international trade data for 3,124 products and 247 countries over the period 1995-2019 and production and domestic trade data for the same number of products and years for 35 European economies. After describing the methods that we employed to construct the GRANTPA database, we demonstrated its usefulness with a gravity application that delivers estimates of several standard gravity variables. We draw two main conclusions about the usefulness of GRANTPA based on this gravity analysis. First, the average estimates that we obtain on each of the standard gravity variables in our econometric model are comparable to the gravity estimates from the existing literature. This reveals that the GRANTPA database is representative in the sense that it captures and reflects the gravity forces that have already been established to shape international (and domestic) trade flows. An alternative interpretation is that gravity works at the very disaggregated level. Second, the disaggregated estimates of all gravity variables in our model vary widely across the products in the GRANTPA database. Consistent with the main motivation for constructing the GRANTPA database, the implication for our database is that more aggregated gravity analysis masks significant heterogeneity, which may be very important from a policy perspective. Accordingly, we expect that the GRANTPA database will be useful for analyzing the effects of various bilateral and country-specific policies.

The norm is that trade theory and trade policy are done in a general equilibrium (GE), e.g., a bilateral free trade agreement or a tariff war between two countries, which may have significant implications for other countries that are not part of the agreement or the tariff war. Proper GE analysis requires consistent trade and production data, and we are aware of some excellent databases that can be used for GE analysis, e.g., Timmer et al. (2015) (WIOD), OECD (2023) (ICIO), and Aguiar et al. (2019) (GTAP) database. However, all existing GE datasets are relatively aggregated (e.g., covering around 50 sectors). As demonstrated, the GRANTPA database can be used to obtain product-level estimates. In terms of GE analysis, we are aware that the GRANTPA database only covers a limited number of countries and that the data is heavily unbalanced. Hence, for future research, we may harmonize and expand the non-EU countries' trade and production data and expand the scope of the database.

References

Aguiar, Angel, Maksym Chepeliev, Erwin L. Corong, Robert McDougall, and Dominique Van Der Mensbrugghe. 2019. “The GTAP Data Base: Version 10.” *Journal of Global Economic Analysis*, 4(1): 1–27.

Anderson, James E., and Eric van Wincoop. 2003. “Gravity with Gravitas: A Solution to the Border Puzzle.” *American Economic Review*, 93(1): 170–192.

Anderson, James E., and Eric van Wincoop. 2004. “Trade Costs.” *Journal of Economic Literature*, 42(3): 691–751.

Arkolakis, Costas, Arnaud Costinot, and Andrés Rodríguez-Clare. 2012. “New Trade Models, Same Old Gains?” *American Economic Review*, 102(1): 94–130.

Bernard, Andrew B, Emily J. Blanchard, Ilke Van Beveren, and Hylke Vandenbussche. 2018. “Carry-Along Trade.” *The Review of Economic Studies*, 86(2): 526–563.

Beverelli, Cosimo, Alexander Keck, Mario Larch, and Yoto V. Yotov. 2023. “Institutions, Trade, and Development: Identifying the Impact of Country-specific Characteristics on International Trade.” *Oxford Economic Papers*.

Correia, Sergio, Paulo Guimarães, and Tom Zylkin. 2020. “Fast Poisson estimation with high-dimensional fixed effects.” *Stata Journal*, 20(1): 95–115.

Costinot, Arnaud, Dave Donaldson, and Ivana Komunjer. 2012. “What Goods Do Countries Trade? A Quantitative Exploration of Ricardo’s Ideas.” *Review of Economic Studies*, 79(2): 581–608.

Gurevich, Tamara, and Peter Herman. 2018. “The Dynamic Gravity Dataset: 1948-2016.” USITC Working Paper 2018-02-A.

Gurevich, Tamara, Peter Herman, Farid Toubal, and Yoto Yotov. 2021. “One Nation, One Language? Domestic Language Diversity, Trade and Welfare.” *CESifo Working Paper*.

Gurevich, Tamara, Peter Herman, Farid Toubal, and Yoto Yotov. 2023. “The Domestic and International Common Language Database (DICL).” Unpublished Manuscript.

Heid, Benedikt, Mario Larch, and Yoto V. Yotov. 2021. “Estimating the Effects of Non-discriminatory Trade Policies within Structural Gravity Models.” *Canadian Journal of Economics*, 54(1): 376–409.

OECD. 2023. “OECD Inter-Country Input-Output Database.” Accessed: December 13 2023.

Pierce, Justin R., and Peter K. Schott. 2012*a*. “A Concordance Between Ten-digit US Harmonized System Codes and SIC/NAICS Product Classes and Industries.” *Journal of Economic and Social Measurement*, 37(1-2): 61–96.

Pierce, Justin R., and Peter K. Schott. 2012*b*. “Concording US Harmonized System Categories over time.” *Journal of Official Statistics*, 28(1): 53–68.

Santos Silva, João M.C., and Silvana Tenreyro. 2006. “The Log of Gravity.” *Review of Economics and Statistics*, 88(4): 641–658.

Shikher, Serge. 2011. “Capital, Technology, and Specialization in the Neoclassical Model.” *Journal of International Economics*, 83(2): 229–242.

Timmer, Marcel P., Erik Dietzenbacher, Bart Los, Robert Stehrer, and Gaaitzen J. De Vries. 2015. “An Illustrated User Guide to the World Input–Output Database: The Case of Global Automotive Production.” *Review of International Economics*, 23(3): 575–605.

Van Beveren, Ilke, Andrew B. Bernard, and Hylke Vandenbussche. 2012. “Concording EU Trade and Production Data over Time.” *National Bureau of Economic Research*, Working Paper (18604).

Yotov, Yoto V. 2022. “On the Role of Domestic Trade Flows for Estimating the Gravity Model of Trade.” *Contemporary Economic Policy*.

Yotov, Yoto V., Roberta Piermartini, José-Antonio Monteiro, and Mario Larch. 2016. *An Advanced Guide to Trade Policy Analysis: The Structural Gravity Model.* Geneva, Switzerland:United Nations and World Trade Organization, available for download at https://www.wto.org/english/res_e/publications_e/advancedguide2016_e.htm.

Appendices

Appendix A Tables and Figures

Table 1: Structure of the Combined Nomenclature (CN8) Classification (Extended)

Combined Nomenclature 8-digit(CN8)		*Harmonized System 6-digit (HS6)*
Year	*# of CN8 products*	
1988	9506	HS6 1988 (# HS6 = 5019)
1989	9579	
1990	9695	
1991	9743	
1992	9837	HS6 1992 (# HS6 = 5018)
1993	9906	
1994	10108	
1995	10448	
1996	10495	HS6 1996 (# HS6 = 5113)
1997	10606	
1998	10587	
1999	10428	
2000	10314	
2001	10274	
2002	10400	HS6 2002 (# HS6 = 5224)
2003	10404	
2004	10174	
2005	10096	
2006	9841	
2007	9720	HS6 2007 (# HS6 = 5051)
2008	9699	
2009	9569	
2010	9443	
2011	9294	
2012	9383	HS6 2012 (# HS6 = 5205)
2013	9376	
2014	9379	
2015	9386	
2016	9414	
2017	9528	HS6 2017 (# HS6 = 5387)
2018	9533	
2019	9533	
2020	9483	
2021	9494	
2022	9736	HS6 2022 (# HS6 = 5612)

Note: All classification files are obtained from Eurostat Ramon server.

Table 2: Changes in the Combined Nomenclature Classification Over Time: Extension

Effective year	Number of obsolete codes	Number of new codes	Number of families (including simple changes)	Number of simple (one-to-one) changes
1989	76	149	58	1
1990	122	238	111	11
1991	85	133	64	8
1992	128	222	85	2
1993	276	345	171	14
1994	233	435	197	11
1995	531	871	383	31
1996	1257	1304	792	435
1997	170	281	130	0
1998	334	315	175	0
1999	303	144	132	3
2000	223	109	96	0
2001	90	50	42	1
2002	847	973	504	311
2003	16	20	12	0
2004	503	273	211	7
2005	186	108	95	5
2006	743	489	281	11
2007	1202	108	630	387
2008	96	75	54	2
2009	257	127	111	0
2010	381	255	151	1
2011	282	133	124	0
2012	959	1048	637	357
2013	43	36	24	1
2014	40	43	22	2
2015	18	25	11	0
2016	27	55	18	0
2017	766	876	414	133
2018	13	18	9	0
2019	9	9	4	0
2020	104	54	42	1
2021	9	20	9	0
2022	535	769	332	135

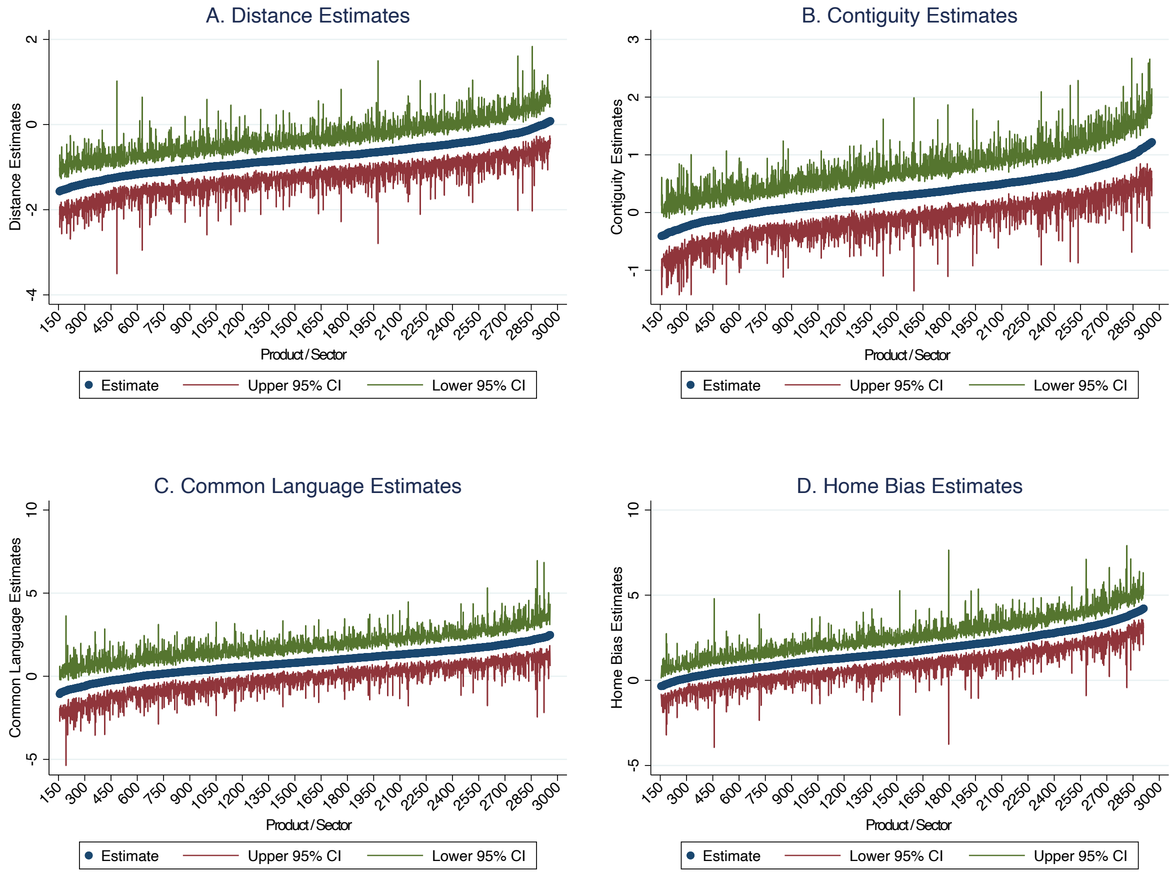
Note: This table shows the number of obsolete and new codes for each year, as well as the number of families (shrinking, growing, or simple) and the number of simple changes (one-to-one). The effective year refers to the year in which the change becomes effective. HS6 codes have been revised in 1992, 1996, 2002, 2007, 2012, 2017 and 2022. The main changes in the combined nomenclature (CN8) classification over time are obtained from Eurostat Ramon server as shown in Van Beveren, Bernard and Vandenbussche (2012).

Table 3: Changes in the Prodcom Classification Over Time: Extension

Effective year	Number of obsolete codes	Number of new codes	Number of families (including simple changes)	Number of simple (one-to-one) changes	Number of codes that are dropped (exit)	Number of codes that are new on the list (entry)
1994	32	46	29	17	4	3
1995	33	52	15	12	19	29
1996	118	80	54	12	14	15
1997	0	0	0	0	0	0
1998	2	0	1	0	2	0
1999	68	92	31	2	3	62
2000	16	12	9	1	0	0
2001	113	76	57	0	0	0
2002	82	54	29	3	1	3
2003	363	296	215	190	1	13
2004	35	24	17	1	1	2
2005	305	105	91	0	67	1
2006	4	2	2	0	0	0
2007	184	131	76	13	3	9
2008	4396	3864	3651	3258	52	19
2009	28	15	15	1	1	1
2010	45	26	23	4	0	0
2011	61	28	28	0	0	0
2012	68	53	40	11	0	5
2013	11	8	1	0	11	8
2014	4	2	1	0	4	2
2015	9	6	1	0	9	6
2016	141	135	95	72	23	28
2017	105	43	1	0	105	43
2018	0	0	0	0	0	0
2019	79	217	1	0	79	217
2020	0	0	0	0	0	0
2021	5	12	1	0	5	12

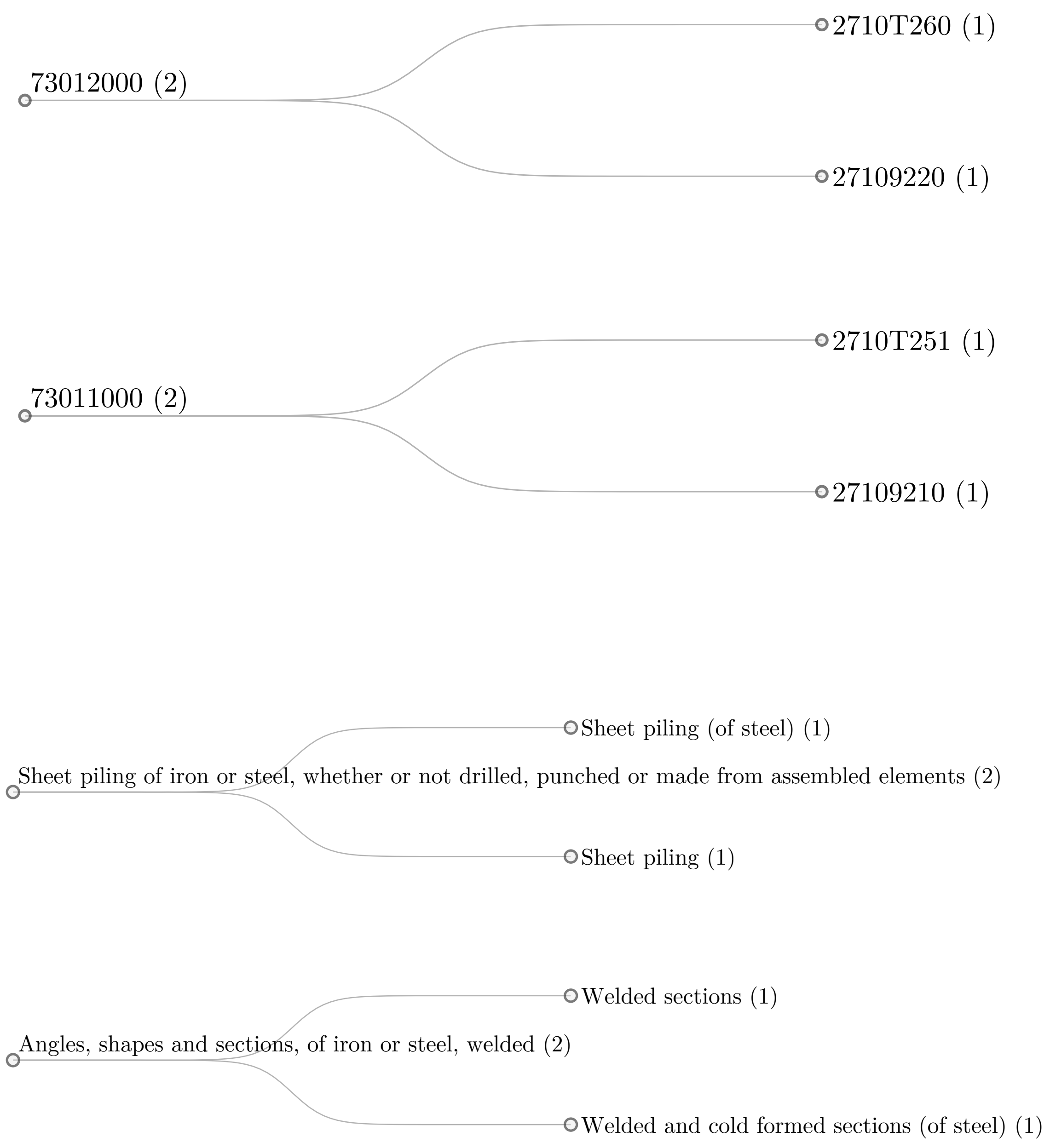
Note: This table shows the number of obsolete and new codes in each year, as well as the number of families (shrinking, growing, simple, entry or exit) and the number of simple changes (one-to-one). The effective year refers to the year in which the change became effective. Some PC8 codes are not covered throughout the whole sample period, resulting in new codes (*entry*) appearing on the list and old codes (*exit*) disappearing from the list. All changes in the PC8 classification over time are obtained from Eurostat Ramon server. Following closely Van Beveren, Bernard and Vandenbussche (2012), optional codes have been removed (or replaced by their mandatory aggregates) to ensure comparability over time and across countries.

Figure 1: Gravity Estimates with the GRANTPA Database, 1995-2019



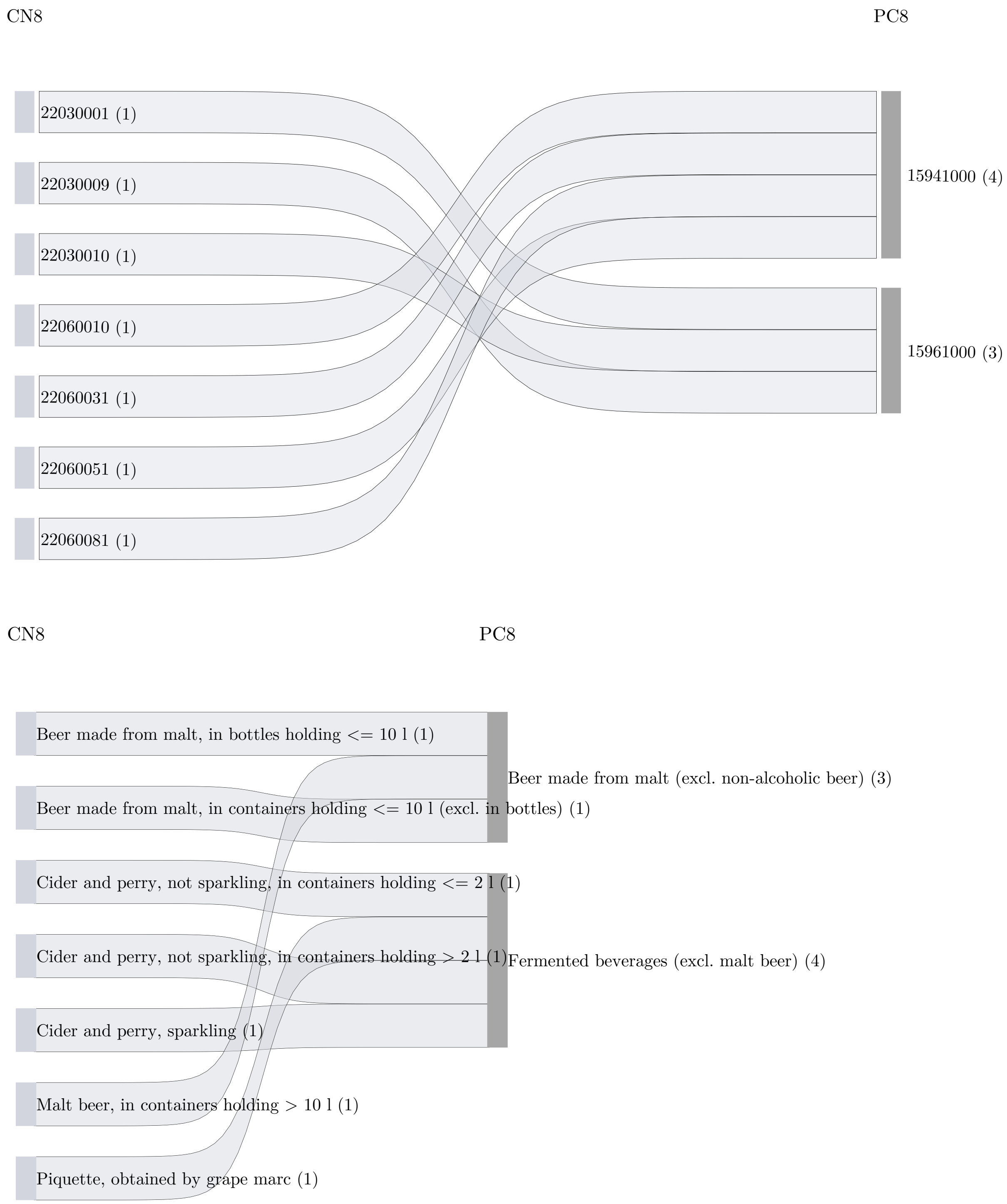
Notes: This figure reports the estimates, along with the corresponding confidence intervals, of the effects on four standard gravity variables. All estimates are obtained with the PPML estimator and exporter-time and importer-time fixed effects according to specification (1), where the dependent variable is always product-level nominal trade in levels from the GRANTPA database. Panel A graphs the estimates of the effects of the log of bilateral distance. Panel B shows the estimates of the effects of contiguous borders. Panel C plots the estimates of the effects of common language. Finally, Panel D visualizes the estimates of the 'home bias'. See text for further details.

Figure 2: Example Complex Codes in 2003: One-to-Many



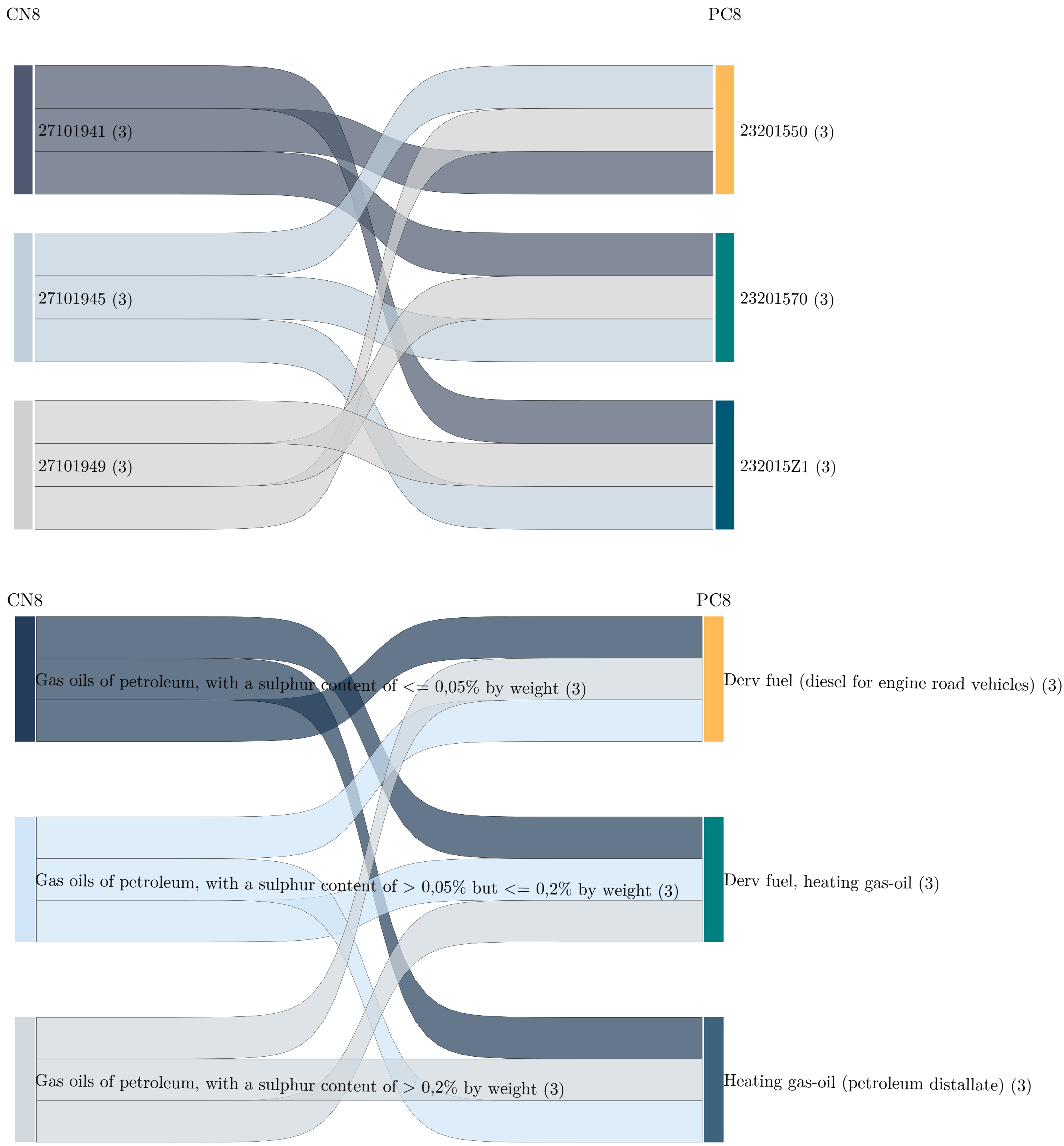
Notes: Mapping of Combined Nomenclature (CN8) to *Prodcom* Codes (PC8). The 8-digit CN code 73012000 (left) is shown mapping to multiple 8-digit *Prodcom codes* 2710T260 and 27109220 (right), demonstrating a one-to-many relationship. Each line represents the transition from a single CN8 code to its corresponding PC8 variants. The numbers in parentheses indicate the frequency relevant to the mapping.

Figure 3: Example Complex Codes in 2003: Many-to-One



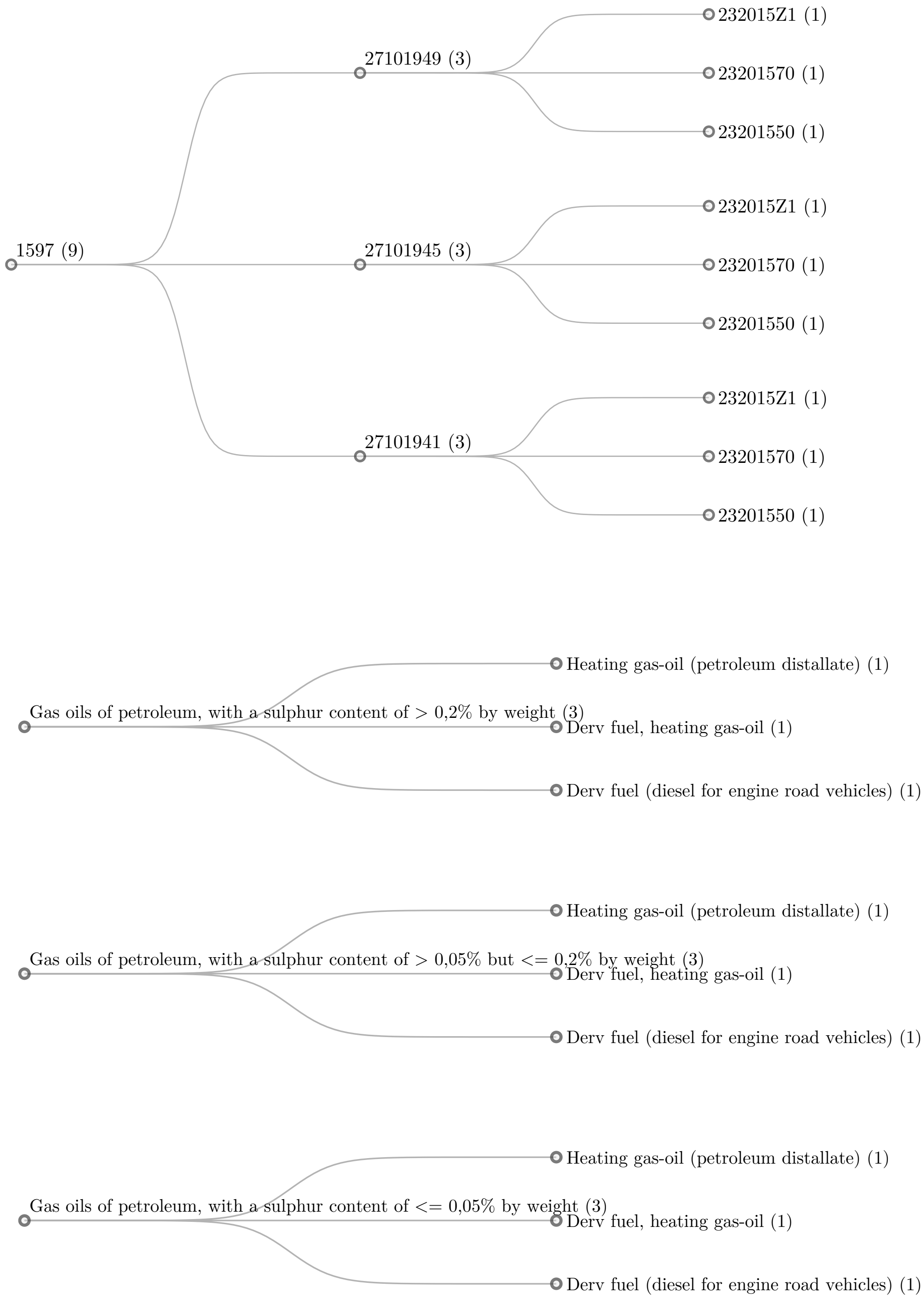
Notes: Mapping of Combined Nomenclature (CN8) (left) to Prodcom codes (PC8) (right). The figure shows a many-to-one relationship, where multiple CN8 codes (22030001, 22030009, 22030010) are linked to one PC8 code 15961000. The lines indicate the directional flow from CN8 to PC8 codes, with the numbers in parentheses denoting the frequency of the mapping process.

Figure 4: Example Complex Codes in 2003: Many-to-Many



Notes: Mapping of Combined Nomenclature (CN8) (left) to Prodcom codes (PC8) (right). This figure presents multiple CN8 codes (left) and their complex mappings to several PC8 codes (right). Each coloured pathway represents the interconnections between a CN8 code and its multiple PC8 counterparts. The numbers in parentheses denote the frequency of the mapping process.

Figure 5: Example Synthetic Codes in 2003: Many-to-Many



Notes: Complex Mapping from '*synthetic*' code (1597) through Combined Nomenclature (CN8) to Prodcom Codes (PC8). This figure illustrates the mapping process starting with PC8+ codes (left), linking through CN8 codes (center), and PC8 codes (right). The numbers in parentheses indicate the frequency.

Figure 6: Growing Family Tree

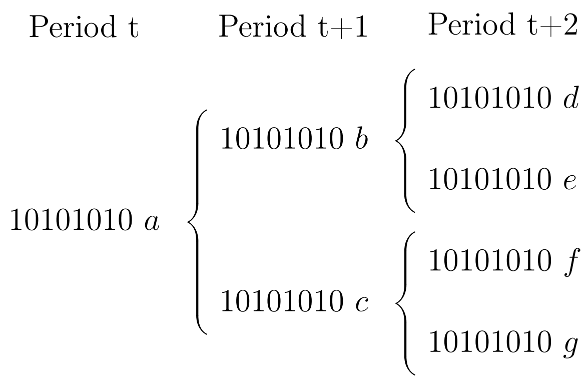
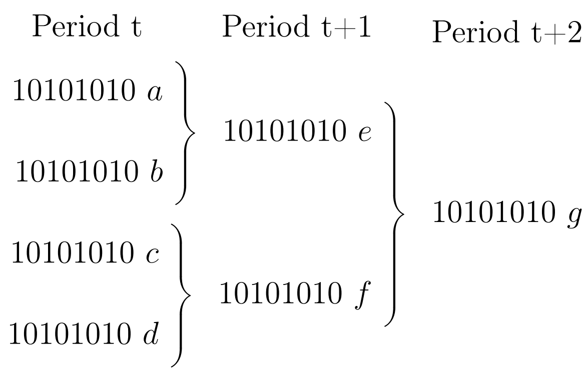


Figure 7: Shrinking Family Tree



IMPRESSUM

Herausgeber, Verleger, Eigentümer und Hersteller:
Verein „Wiener Institut für Internationale Wirtschaftsvergleiche“ (wiiw),
Wien 6, Rahlgasse 3

ZVR-Zahl: 329995655

Postanschrift: A 1060 Wien, Rahlgasse 3, Tel: [+431] 533 66 10, Telefax: [+431] 533 66 10 50
Internet Homepage: www.wiiw.ac.at

Offenlegung nach § 25 Mediengesetz: Medieninhaber (Verleger): Verein "Wiener Institut für Internationale Wirtschaftsvergleiche", A 1060 Wien, Rahlgasse 3. Vereinszweck: Analyse der wirtschaftlichen Entwicklung der zentral- und osteuropäischen Länder sowie anderer Transformationswirtschaften sowohl mittels empirischer als auch theoretischer Studien und ihre Veröffentlichung; Erbringung von Beratungsleistungen für Regierungs- und Verwaltungsstellen, Firmen und Institutionen.